Quantum partial information decomposition

S. J. van Enk
Department of Physics and Oregon Center for Optical, Molecular & Quantum Sciences, University of Oregon, Eugene, Oregon 97403, USA

(Received 8 August 2023; accepted 13 November 2023; published 13 December 2023)

Partial information decomposition (PID) takes one step beyond Shannon's theory in decomposing the information two variables, A and B, possess about a third variable, $\mathcal{T}$, into distinct parts: unique, shared (or redundant), and synergistic information. Here we show how these concepts can be defined in a quantum setting. We apply a quantum PID to scrambling in quantum many-body systems, for which a quantum-theoretic description has been proven to be productive. Unique information in particular provides a finer description of scrambling than does the so-called tri-information.

DOI: 10.1103/PhysRevA.108.062415

I. INTRODUCTION

Partial information decomposition (PID) is a method for disentangling the relations between multiple random variables as encoded in their joint probability distribution. The method was conceived in Ref. [1] and, in the simplest nontrivial case of three variables [2], defines synergistic information, unique information, and redundant information. One important motivating example for desiring to go beyond Shannon's information theory is given in Ref. [3]. There two different probability distributions P_1 and P_2 over three variables with distinct underlying mechanisms which cannot be distinguished by any of the standard quantities defined within Shannon's theory were presented. Specifically, mutual information between any combination of the three variables cannot distinguish P_1 from P_2. Hence, linear combinations of mutual information quantities, such as coinformation [4] (whose negative is known in the quantum context as tri-information), also cannot distinguish P_1 and P_2. The PID, however, does distinguish them: all newly introduced quantities—synergistic, unique, and shared information—differ for P_1 and P_2. (See Sec. II for the details.)

Using its ability to make distinctions between different probabilistic mechanisms, the PID has been applied to descriptions and understanding of complex networks [5,6] and neural networks in particular [7–10]. For an overview of its uses, see [11]. Several different proposals for PID exist based on different definitions of either unique information [12,13] or redundant information [1,14–18] or synergistic information [19,20]. The idea that a finer distinction between different types of information is useful was around in neuroscience [21–23] before the seminal work in Ref. [1].

Since all of Shannon's classical concepts have been generalized to quantum settings and since such generalizations have proven very fruitful [24–27], there ought to be a quantum version of the PID (QPID) as well. We define our version in Sec. III. As we will show in Sec. IV, the classical motivating example distributions P_1 and P_2 can be quantized such that standard quantum mutual information quantities (including tri-information) do not distinguish the two corresponding pure states $|\Psi_1\rangle$ and $|\Psi_2\rangle$. Just like the classical PID, the QPID is not unique, but the version proposed here does clearly distinguish these two pure states. Moreover, when applying the PID concepts to the issue of quantum scrambling [28] in quantum many-body systems, the particular choice we make here gives sensible numerical results, especially for the unique information. There is good reason to expect unique information to play a role in a quantum context. Whereas classical correlations can be shared unrestrictedly, the no-cloning theorem (or the monogamy of entanglement; see, e.g., Ref. [29]) prohibits two systems from possessing the same maximal quantum entanglement as a third system, thus typically forcing both systems to possess some unique quantum information about the third system. (See Sec. IV B for details.)

We start by summarizing some relevant aspects of the classical PID.

II. PID FOR THREE VARIABLES

Consider two classical variables A and B correlated with a third classical variable $\mathcal{T}$ that is the target of our inquiry. We assume the joint probability distribution $P(\mathcal{T}, A, B)$ exists. We may then define the Shannon entropy [30,31]

$$H(X) = -\sum_{\mathcal{T},A,B} P(\mathcal{T}, A, B)\log_2[P(X)], \tag{1}$$

in which the symbol X may stand for any subset of the variables A, B, and $\mathcal{T}$ as well as for any conditional variable, such as $\mathcal{T}|A$. We use here logarithms in base 2, so information and entropy will be given in units of bits.

The amount of information about $\mathcal{T}$ we can obtain from variable A alone is given by the mutual information [30,31]

$$I(\mathcal{T};A) = H(\mathcal{T}) - H(\mathcal{T}|A) \tag{2a}$$

$$= H(\mathcal{T}) + H(A) - H(\mathcal{T}, A). \tag{2b}$$

That is, according to the first line, information equals the decrease in entropy when going from an initial distribution $P(\mathcal{T})$ to a final probability distribution $P(\mathcal{T}|A)$. The second line shows the mutual information thus defined is actually symmetric between the two variables A and $\mathcal{T}$.

TABLE I. Probability distribution taken from Ref. [3] (the "triadic" case). Each of the three variables can take on four values, 0, 1, 2, or 3. The information obtainable from $P(\mathcal{T}|A)$ is the same as that obtainable from $P(\mathcal{T}|B)$. Namely, when A is even (odd), we can conclude $\mathcal{T}$ is even (odd), but nothing more, and the same conclusion about $\mathcal{T}$ follows from B.

$\mathcal{T}, A, B$	Probability
0,0,0	1/8
0,2,2	1/8
2,0,2	1/8
2,2,0	1/8
1,1,1	1/8
1,3,3	1/8
3,1,3	1/8
3,3,1	1/8

An important motivation for extending Shannon's theory and going beyond mutual information is the following example, taken from Ref. [3]. In Tables I and II we display two different joint probability distributions for our three variables, $\mathcal{T}, A$, and B, each taking on four possible values. These two distributions cannot be distinguished by any measures constructed from just the Shannon entropies of the above type.

For example, each has a joint entropy $H(\mathcal{T}, A, B) = 3$, and for each we have $I(\mathcal{T}; A) = I(\mathcal{T}; B) = 1$. Using more general entropy functions, such as the Rényi entropy H_α for $\alpha \geqslant 0$ [32], does not help either in distinguishing the two [3]. As captions of Tables I and II explain, there is a difference, however, in what sort of information the individual variables A and B carry about $\mathcal{T}$. Even though the amounts are the same (1 bit), in the first case the variables carry the same bit of information, and in the second they carry different bits. In other words, in the latter case, each variable carries some information that is unique with respect to the other variable.

We thus wish to quantify how much information about $\mathcal{T}$ from B is unique (relative to A, that is) and how much information is unique to A (relative to B). In the context of the PID we frame this problem as follows. We try to write the three known standard mutual information quantities $I(\mathcal{T}; A)$, $I(\mathcal{T}; B)$, and $I(\mathcal{T}; A, B)$ that concern information about $\mathcal{T}$ as linear combinations of four new quantities (only one of which is therefore linearly independent of the three standard quantities). Two of these new quantities are meant to quantify the two types of unique information; the other two then give redundant and synergistic information as follows:

$$\begin{aligned} I(\mathcal{T}; A) &= I_{\text{unq}\,A\setminus B} + I_{\text{red}\,A\&B}, \\ I(\mathcal{T}; B) &= I_{\text{unq}\,B\setminus A} + I_{\text{red}\,A\&B}, \\ I(\mathcal{T}; A, B) &= I_{\text{unq}\,A\setminus B} + I_{\text{unq}\,B\setminus A} + I_{\text{red}\,A\&B} + I_{\text{syn}\,A\&B}. \end{aligned} \tag{3}$$

Since we always consider information about the variable $\mathcal{T}$, we suppress reference to it in the quantities appearing on the right-hand side.

Since we have introduced four quantities but have only three equations for them, one more constraint is needed to define the PID. Disagreement has arisen in the literature over what the fourth equation ought to be. Our view is that there is, indeed, a freedom of choice and that distinct choices make sense in different contexts. Here we will make a choice that can be straightforwardly generalized to a quantum setting (but there may well be other such choices [33]).

Here is one way to define unique information [20]. We define for each pair of possible values a and b for the variables A and B

$$Z_{ab} = \sum_t P(t|a)^{1/2} P(t|b)^{1/2}, \tag{4}$$

where the sum is over the values t that the variable $\mathcal{T}$ can take. This is an overlap between two distribution functions, the Bhattacharyya measure [34]. It lies between 0 (for orthogonal distributions, which have no common support; this extreme case cannot occur in our setting) and 1 (for identical distributions). The idea is that unique information exists only when the conditional distributions $P(t|a)$ and $P(t|b)$ are different for at least some possible values of a and b. We define the non-negative quantity $\mathcal{B}_1$ by

$$\mathcal{B}_1 = -\sum_a \sum_b P(a, b) \log_2(Z_{ab}), \tag{5}$$

where

$$P(a, b) = \sum_t P(t, a, b) \tag{6}$$

is the joint distribution for a and b. An operational meaning for $\mathcal{B}_1$ is given in Ref. [20] in terms of pooling probability distributions [35–38]: given the two distributions $P(t|a)$ and $P(t|b)$ for fixed a and b, we can generate a single distribution over t by choosing

$$P_{\text{pool}}(t) = P(t|a)^{1/2} P(t|b)^{1/2} / Z_{ab}, \tag{7}$$

where Z_{ab} appears as a normalization factor. (This method of pooling is called "logarithmic.") For this pooled distribution the average uncertainty about $\mathcal{T}$ is reduced relative to the average entropy $\frac{1}{2}[H(\mathcal{T}|A) + H(\mathcal{T}|B)]$ by an amount $\mathcal{B}_1$. $\mathcal{B}_1$ can thus be considered a "bonus" amount of information.

An alternative way to define a single distribution with less uncertainty than this average amount is the trivial way

TABLE II. Probability distribution taken from Ref. [3] (the "dyadic" case), which should be contrasted with the distribution from Table I. Each of the three variables can take on four values, 0, 1, 2, or 3. The information obtainable from $P(\mathcal{T}|A)$ differs from that obtainable from $P(\mathcal{T}|B)$. For example, from $A = 0$ we can conclude that $\mathcal{T} = 0$ or 1. From $B = 0$ we can conclude $\mathcal{T} = 0$ or 2. There is a crucial bit of information we get uniquely from A and another unique bit we get from B. Together, they fix the value of $\mathcal{T}$ (in this example, we conclude $\mathcal{T} = 0$).

$\mathcal{T}, A, B$	Probability
0,0,0	1/8
0,2,1	1/8
1,0,2	1/8
1,2,3	1/8
2,1,0	1/8
2,3,1	1/8
3,1,2	1/8
3,3,3	1/8

of choosing whichever of $P(t|a)$ and $P(t|b)$ has the lowest entropy on average (averaged over all values of a and b). For that trivial choice the reduction in uncertainty (an alternative "bonus") would equal

$$\mathcal{B}_0 = \tfrac{1}{2}|H(\mathcal{T}|A) - H(\mathcal{T}|B)|. \tag{8}$$

We then define unique information by using the larger of the bonuses $\mathcal{B}_1$ and $\mathcal{B}_0$,

$$\mathcal{B} := \max(\mathcal{B}_0, \mathcal{B}_1), \tag{9}$$

as

$$I_{\text{unq}\,A\setminus B} + I_{\text{unq}\,B\setminus A} = 2\mathcal{B}. \tag{10}$$

The individual unique information quantities then follow from (3),

$$\begin{aligned} I_{\text{unq}\,A\setminus B} &= \mathcal{B} + \tfrac{1}{2}[H(\mathcal{T}|B) - H(\mathcal{T}|A)], \\ I_{\text{unq}\,B\setminus A} &= \mathcal{B} + \tfrac{1}{2}[H(\mathcal{T}|A) - H(\mathcal{T}|B)], \end{aligned} \tag{11}$$

which are both non-negative. We note that with these definitions we can, indeed, clearly distinguish the two probability distributions P_1 and P_2 from Tables I and II: we find $\mathcal{B} = 0$ for P_1 (and hence zero unique information) but $\mathcal{B} = \mathcal{B}_1 = 1$ (and hence 1 bit of unique information for both A and B) for P_2, exactly agreeing with the intuition given in the captions.

We may view $\mathcal{B}$ as quantifying the asymmetry between A and B with respect to their correlations with $\mathcal{T}$. We may also view $\mathcal{B}$ as the fourth independent quantity, in addition to the three mutual information measures, that characterizes how two variables may contain information about a third variable.

To conclude this section on the classical PID, let us note that not all quantities named "information" in the decomposition (3) are differences between two entropies. In particular, for the decomposition based on (11) and, similarly, for the decomposition proposed in [12], the "redundant information" actually is a difference between two information quantities. As such, it was denoted by $\Delta I_{\text{red}\,A\&B}$ in early works [22,23], as well as in [20]. Being a difference between two (non-negative) information quantities, $\Delta I_{\text{red}\,A\&B}$ may take on negative values. It follows directly from (3) that unique information is larger than the mutual information in such a case. The same will then be true of its quantum generalization.

III. FROM CLASSICAL TO QUANTUM INFORMATION

When trying to define quantum information quantities by replacing classical variables by quantum systems, replacing probability distributions by density operators ρ, and replacing the Shannon entropy function by the von Neumann entropy $S(\rho) = -\text{Tr}\rho\log_2(\rho)$, one runs into two well-known sorts of issues. First, different density operators do not commute in general. Second, unlike in the classical case, whether we perform measurements or not and, in addition, what measurements we perform matter. Both issues are illustrated below.

Quantum information theory [24–27] has taught us how to circumvent such difficulties and define meaningful quantities that generalize classical quantities. Often, multiple generalizations exist for a given classical quantity (with different interpretations) for the reasons alluded to above. We will see that here, too, multiple possibilities exist to generalize the PID to a quantum version. For various reasons mentioned below, we do focus here on one particular generalization.

A. Mutual information

Given the two equivalent ways [(2a) and (2b)] in which we can define classical mutual information $I(\mathcal{T};A)$, there are two ways to generalize mutual information to a quantum setting. As is well known, these two generalizations are no longer equivalent.

The first generalization is explicitly not symmetric between the two variables: we assume that we perform a measurement on system A and produce a density operator for system $\mathcal{T}$ that depends on the measurement outcome. We can always describe the measurement on A by a set of positive operator-valued measure elements $\{\Pi_n\}$ labeled by the outcome n. Suppose the state of $\mathcal{T}$ changes to $\rho_{\mathcal{T}|n}$ if outcome n occurs; this occurs with probability $p_n = \text{Tr}(\rho_{\mathcal{T}A}\Pi_n)$. The average von Neumann entropy of system $\mathcal{T}$ after the measurement is then

$$S(\mathcal{T}|A) = \sum_n p_n S(\rho_{\mathcal{T}|n}). \tag{12}$$

In terms of the reduced state for system $\mathcal{T}$ we define

$$J(\mathcal{T};A) = S(\rho_{\mathcal{T}}) - \sum_n p_n S(\rho_{\mathcal{T}|n}). \tag{13}$$

This generalizes (2a). As is fairly standard, we used a different symbol here, J, to denote this particular quantum version of the mutual information [39]. J depends on what measurement is performed on A. We may eliminate this dependence by maximizing J over all possible measurements. This optimization, however, is computationally hard [40].

The symbol I is used for the other quantum definition of mutual information based on the symmetric classical definition (2b). This alternative definition involves no measurements and reads

$$I(\mathcal{T};A) = S(\rho_{\mathcal{T}}) + S(\rho_A) - S(\rho_{\mathcal{T}A}). \tag{14}$$

This is the definition we will use here. We note the well-known property that the mutual information between two qubits can equal 2, namely, when their state is maximally entangled, whereas the classical quantity can be at most 1. This factor of 2 can be interpreted operationally via superdense coding [41], which demonstrates how one qubit of an entangled pair can be used to transmit two classical bits of information.

For the QPID, too, we wish to find expressions that do not involve measurements using expression (4) as a classical starting point. We thus need the quantum version of a conditional state.

B. Quantum conditional states

Classically, we have the conditional probability distribution $P(\mathcal{T}|A) = P(\mathcal{T},A)/P(A)$. We cannot straightforwardly find a quantum equivalent to this distribution because the two operators $\rho_{\mathcal{T}A}$ and $\mathbb{1}_{\mathcal{T}} \otimes \rho_A$ do not necessarily commute. Indeed, there cannot be a conditional state $\rho_{\mathcal{T}|A}$ such that we

always have

$$S(\rho_{\mathcal{T}|A}) = S(\rho_A) - S(\rho_{\mathcal{T}A}) \tag{15}$$

simply because the quantity on the right-hand side may be negative. If it is negative, $\rho_{\mathcal{T}A}$ is entangled.

Nonetheless, we can define an operator that generalizes $P(\mathcal{T}|A) = P(\mathcal{T}, A)/P(A)$ and that has at least some properties in common with a conditional probability distribution. Here we choose to define the quantum conditional state as

$$\rho_{\mathcal{T}|A} = \left(\mathbb{1}_{\mathcal{T}} \otimes \rho_A^{-1/2}\right) \rho_{\mathcal{T}A} \left(\mathbb{1}_{\mathcal{T}} \otimes \rho_A^{-1/2}\right). \tag{16}$$

This is an operator on the joint system $\mathcal{T} \otimes A$ but fails to be a density operator in that it may have eigenvalues larger than 1 (and hence its entropy may be negative). This operator can be used to witness bipartite entanglement [42]. The symbol $\rho_A^{-1/2}$ here is defined straightforwardly with the support of ρ_A and is set to zero on its kernel (thus corresponding to the Moore-Penrose inverse). One useful perspective on this quantum version of the conditional probability distribution can be found in Refs. [43,44], where it is argued that this is the proper quantum generalization from a Bayesian perspective. References [43,44] also demonstrate the usefulness of the $*$ product, defined as

$$A * B := B^{1/2} A B^{1/2}, \tag{17}$$

such that $\rho_{\mathcal{T}|A} = \rho_{\mathcal{T}A} * (\rho_A)^{-1}$, where additional identity operators have been suppressed.

An alternative definition of conditional states was proposed in Refs. [45,46],

$$\rho'_{\mathcal{T}|A} = \lim_{n\to\infty} \left[\mathbb{1}_{\mathcal{T}} \otimes (\rho_A)^{-1/(2n)} \, \rho_{\mathcal{T}A}^{1/n} \, (\mathbb{1}_{\mathcal{T}} \otimes \rho_A)^{-1/(2n)}\right]^n. \tag{18}$$

It can likewise be employed to detect negative conditional entropy [47] and entanglement [48]. As a test of entanglement it is weaker than that based on (16). This is one reason for using the conditional operator as defined in (16) in our definition of a QPID. The other reason is that the relation between $\rho_{\mathcal{T}|A}$ and $\rho_{\mathcal{T}A}$ in (16) can be inverted straightforwardly (in fact, by using the $*$ product) but (18) cannot. We thus use definition (16).

IV. QUANTUM PID

We wish to define quantum versions of Z_{ab} as given in (4) and of $\mathcal{B}_1$ as given in (5). We choose here

$$Z_{AB} = \tfrac{1}{2}\mathrm{Tr}_{\mathcal{T}}\left(\rho_{\mathcal{T}|A}^{1/4} \rho_{\mathcal{T}|B}^{1/2} \rho_{\mathcal{T}|A}^{1/4} + \rho_{\mathcal{T}|B}^{1/4} \rho_{\mathcal{T}|A}^{1/2} \rho_{\mathcal{T}|B}^{1/4}\right). \tag{19}$$

Here, of course, identities on A and B have to be inserted in the definitions of $\rho_{\mathcal{T}|B}$ and $\rho_{\mathcal{T}|A}$, respectively, to produce operators on $A \otimes B \otimes \mathcal{T}$. In terms of the $*$ product we may rewrite this as

$$Z_{AB} = \tfrac{1}{2}\mathrm{Tr}_{\mathcal{T}}\left(\rho_{\mathcal{T}|B}^{1/2} * \rho_{\mathcal{T}|A}^{1/2} + \rho_{\mathcal{T}|A}^{1/2} * \rho_{\mathcal{T}|B}^{1/2}\right). \tag{20}$$

The operator Z_{AB} is then used to define

$$\mathcal{B}_{\mathcal{Q}1} = -\mathrm{Tr}_{AB}[\rho_{AB} \log_2(Z_{AB})] \tag{21}$$

by analogy to the classical quantity $\mathcal{B}_1$. The analog of $\mathcal{B}_0$ is then

$$\mathcal{B}_{\mathcal{Q}0} = \tfrac{1}{2}|I(\mathcal{T};A) - I(\mathcal{T};B)|, \tag{22}$$

and we replace (9) by

$$\mathcal{B}_{\mathcal{Q}} = \max(\mathcal{B}_{\mathcal{Q}0}, \mathcal{B}_{\mathcal{Q}1}). \tag{23}$$

[In the Appendix we numerically compare these two pooling methods for typical quantum states of three qubits and of three qutrits, with the result that typically (but not always), the more complicated logarithmic pooling method (yielding $\mathcal{B}_{\mathcal{Q}1}$) is superior to the trivial method (yielding $\mathcal{B}_{\mathcal{Q}0}$).]

We define (non-negative) unique information by analogy to (11),

$$\begin{aligned} I_{\mathrm{unq}\,A\setminus B} &= \mathcal{B}_{\mathcal{Q}} + \tfrac{1}{2}[I(\mathcal{T};A) - I(\mathcal{T};B)], \\ I_{\mathrm{unq}\,B\setminus A} &= \mathcal{B}_{\mathcal{Q}} + \tfrac{1}{2}[I(\mathcal{T};B) - I(\mathcal{T};A)]. \end{aligned} \tag{24}$$

Note that we might alternatively use

$$Z'_{AB} = \tfrac{1}{2}\mathrm{Tr}_{\mathcal{T}}\left(\rho_{\mathcal{T}|A}^{1/2} \rho_{\mathcal{T}|B}^{1/2} + \rho_{\mathcal{T}|B}^{1/2} \rho_{\mathcal{T}|A}^{1/2}\right), \tag{25}$$

which does not use the $*$ product and thus lacks some of its nice properties. This alternative expression is used for some numerical examples in the next two sections to show that certain features found there are insensitive to this choice.

Finally, note that the definition of $\mathcal{B}_{\mathcal{Q}}$, like that of quantum mutual information, is invariant under local unitary transformations of the form $U = U_T \otimes U_A \otimes U_B$. All information measures introduced here are therefore invariant under local unitaries as well.

A. Motivating example

Let us consider two pure [49] states corresponding to the classical distributions from Tables I and II, defined by taking equal superpositions of the eight possible three-party terms appearing there, and let us denote the two states by $|\Psi_1\rangle$ and $|\Psi_2\rangle$, respectively. We first note that the quantum mutual entropies do not distinguish these two states. For example, for both states we have $S(X) + S(\mathcal{T}) - S(\mathcal{T}, X) = 2$ for $X = A$ and for $X = B$, and $S(A, B) + S(\mathcal{T}) - S(\mathcal{T}, A, B) = 4$. But the quantity $\mathcal{B}_{\mathcal{Q}}$ is clearly different for these two states: it equals 2 for $|\Psi_2\rangle$ and 0 for $|\Psi_1\rangle$. [The same values are found for the alternative definition based on (25).] Thus, the quantity $\mathcal{B}_{\mathcal{Q}}$ straightforwardly generalizes the classical quantity $\mathcal{B}$ in this particular case. The quantity $\mathcal{B}_{\mathcal{Q}}$ can be twice as large as the corresponding classical quantity $\mathcal{B}$, just as the quantum mutual information can be twice as large as the classical mutual information. This indicates $\mathcal{B}_{\mathcal{Q}}$ contains both classical and quantum unique information.

B. Scrambling

The three-variable PID can be applied straightforwardly to the Hayden-Preskill model of a black hole as a processor of quantum information [28]. In their setup, adapted to our notation, $\mathcal{T}$ is a reference system that is initially maximally entangled with a specific small subsystem of $A \otimes B$. That subsystem is thrown into a black hole, and the information (i.e., the entanglement with $\mathcal{T}$) is then scrambled among all degrees of freedom of the full system $A \otimes B$. System B then is emitted by the black hole as Hawking radiation, and Bob, who collects all that radiation, tries to recover the entanglement with $\mathcal{T}$. The part that remains inside the black hole corresponds to system A.

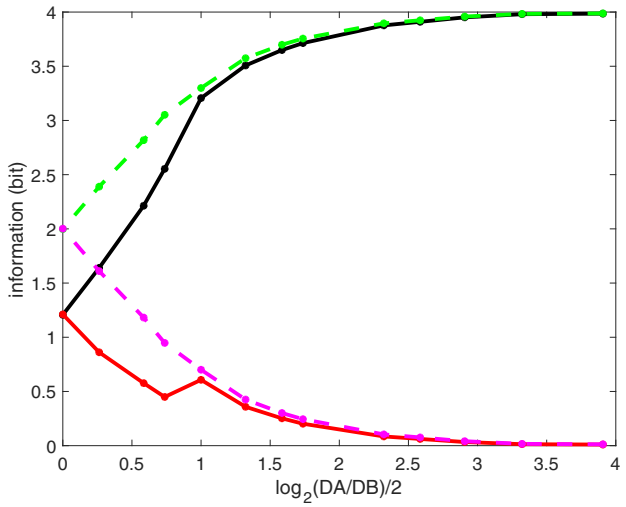

FIG. 1. Result of applying a random unitary to system AB, starting with a maximally entangled state between $\mathcal{T}$ and AB. Keeping the product of Hilbert space dimensions D_A and D_B fixed and equal to 900 [we could view this system as consisting of two qubits, two qutrits, and two five-dimensional systems (ququints)], we plot unique information for A (solid black upper curve) and for B (solid red lower curve), as well as the mutual information $I(\mathcal{T};A)$ (dashed green upper curve) and $I(\mathcal{T};B)$ (dashed magenta lower curve), all as a function of $\log_2(D_A/D_B)/2$. The latter can be interpreted as the effective number of qubits that are still part of system A and not yet part of system B, counted from the starting point with equal dimensions $D_A = D_B = 30$.

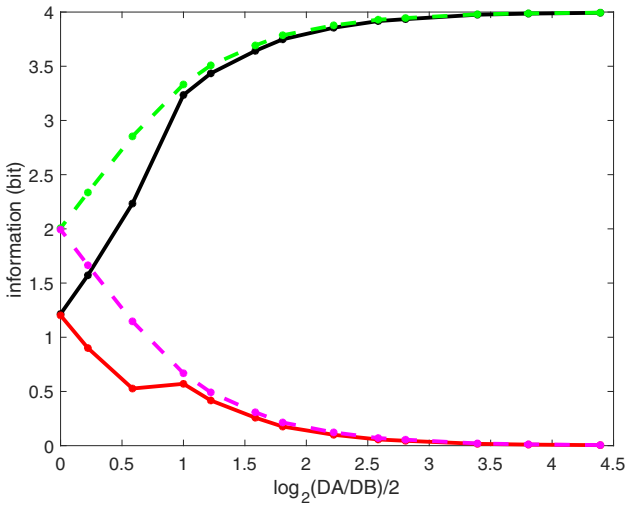

FIG. 2. Same as Fig. 1, but using $D_{AB} = (2 \times 3 \times 7)^2 = 1764$.

In our numerical example we choose the reference system $\mathcal{T}$ to have Hilbert space dimension $D_\mathcal{T} = 4$ and then choose a pure *final* state for $\mathcal{T}AB$ such that $\mathcal{T}$ is (still) maximally entangled with AB. Specifically, we write

$$|\Psi\rangle = \frac{1}{\sqrt{D_\mathcal{T}}} \sum_{n=1}^{D_\mathcal{T}} |n\rangle_\mathcal{T} |\Phi_n\rangle_{AB}, \tag{26}$$

where $\{|\Phi_n\rangle_{AB}\}$ for $n = 1, \ldots, D_\mathcal{T}$ are random orthogonal unit vectors on AB. This final state mimics the result of the scrambling of quantum information, which is initially located in a small subsystem of AB, spreading throughout the entire system AB (as in [28]). In the usual setup [28] AB consists of qubits, and one studies how the information quantities $I(\mathcal{T};A)$ and $I(\mathcal{T};B)$ change when qubits are moved from A to B. The ratio D_A/D_B can change only by factors of 4 in that case. Here we wish to have a more fine-grained description, and we allow qutrits and ququints ($D = 5$) as well, thus making more values for the ratio D_A/D_B possible. We use $D_{AB} = 900 = (2 \times 3 \times 5)^2$ in Fig. 1.

We see in Fig. 1 that the information about $\mathcal{T}$ is equally distributed among A and B when they have the same dimension (as it should be). The green curve gives $I(\mathcal{T};A)$, which quickly approaches the maximum information (4 bits) available in systems A and B together as the dimension of system A becomes larger than half the total size. Two points illustrated in Fig. 1 are worth mentioning. First, when the sizes of A and B are the same, more than half of the information each system possesses is unique. Second, once just one qubit has moved from one system to the other (so that the ratio of their dimensions is 4), almost all information possessed by either system is unique.

Note that the unique information in the smaller subsystem tends to decrease with decreasing size, except around the point where the ratio of two Hilbert space dimensions is almost 4 (i.e., starting from equal dimensions, we have moved one qubit from one system to the other). At that point, swapping a qubit and a qutrit from A for a ququint from B slightly decreases the unique information in B even though its dimension increases. For the larger subsystem the amount of unique information always increases with increasing dimension, but there is an inflection point clearly visible in the black curve when the ratio of the dimensions is 4. We checked this behavior for a dimension of $D_{AB} = 1764$ (see Fig. 2), with the same result. All these conclusions hold as well for the alternative measure based on (25), as shown in Fig. 3.

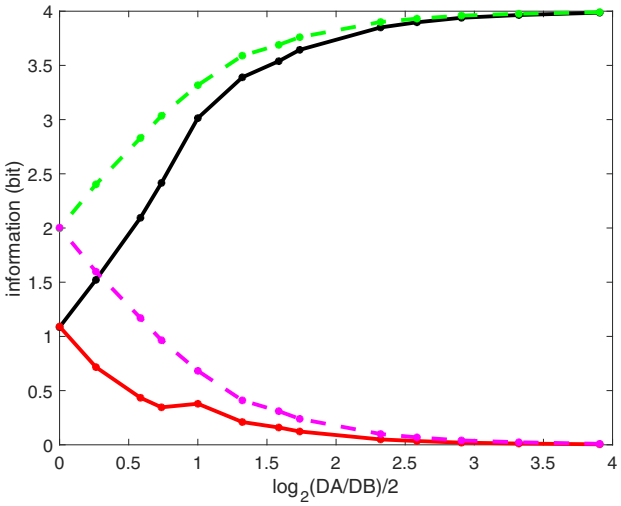

FIG. 3. Same as Fig. 1, but using the alternative expression (25) for Z'_{AB}.

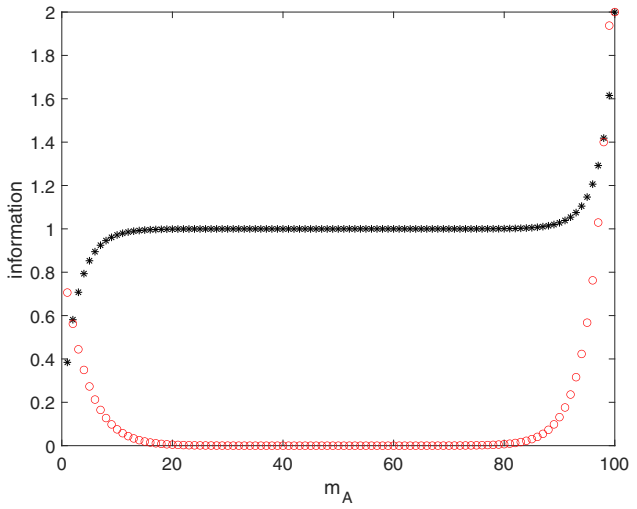

FIG. 4. Mutual information $I(\mathcal{T};A)$ (stars) and unique information $I_{\mathrm{unq}\,A\backslash B}$ (circles) plotted as a function of $m_A = 1,\ldots,N$ for a total of $N = 100$ qubits. The target system is one qubit, system A consists of m_A qubits, and system B consists of $N - m_A$ qubits. Here $s = 0.85$ and $p = q = 1/2$.

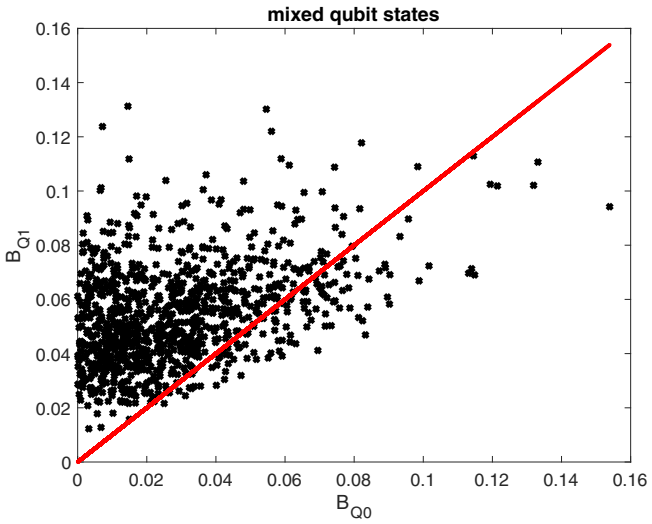

FIG. 5. For 1000 randomly chosen mixed states of three qubits, $\mathcal{B}_{\mathcal{Q}1}$ vs $\mathcal{B}_{\mathcal{Q}0}$, with the red solid line indicating $\mathcal{B}_{\mathcal{Q}1} = \mathcal{B}_{\mathcal{Q}0}$. For a bit more than 11% of the states $\mathcal{B}_{\mathcal{Q}1}$ fails to be larger than $\mathcal{B}_{\mathcal{Q}0}$.

C. Emerging classicality

One outstanding problem of quantum mechanics is explaining how a classical world emerges from a quantum world. A slightly more manageable version of this problem is explaining how classical information arises out of quantum information. In this context, the main point of quantum Darwinism [39,50–53] is that classicality (of a system) emerges from quantum mechanics by having many copies of the same information (about that system) spread out over the environment. Here we consider this same problem in terms of unique information. Let us thus reanalyze an example from Ref. [52]. System $\mathcal{T}$ is one qubit, system A consists of m_A qubits, and system B consists of m_B qubits. The initial state of $\mathcal{T}$ is taken to be $|\psi\rangle_{\mathcal{T}} = \sqrt{p}|0\rangle + \sqrt{q}|1\rangle$, with $p + q = 1$. Each of the $N := m_A + m_B$ qubits starts in state $|0\rangle$ and then interacts with $\mathcal{T}$ by a unitary two-qubit interaction, described as a Controlled-MAYBE (C-MAYBE) in Ref. [52]. This interaction transforms the initial state to a final state of the form

$$|\Psi\rangle_{\mathcal{T},A,B} = \sqrt{p}|0\rangle \otimes |0\rangle^{\otimes m_A} \otimes |0\rangle^{\otimes m_B} + \sqrt{q}|1\rangle \otimes |r\rangle^{\otimes m_A} \otimes |r\rangle^{\otimes m_B}, \tag{27}$$

where $|r\rangle$ is a slightly rotated state,

$$|r\rangle = s|0\rangle + \sqrt{1-s^2}|1\rangle, \tag{28}$$

where s is a real number fairly close to 1. In this way, each qubit of the environment encodes some partial information about the state of qubit $\mathcal{T}$, while all environment qubits taken together contain almost the maximum of 2 bits of quantum mutual information. This simple intuition is quantified in Fig. 1 of Ref. [52] and also in Fig. 4. In both figures there is a clear "classical" plateau visible for which the m_A qubits contain about 1 bit of classical information about qubit $\mathcal{T}$. That is, the quantum mutual information equals about 1 bit for a large range of values of m_A. Using our measure for $I_{\mathrm{unq}\,A\backslash B}$, we also see that that information is not unique to A. Indeed, the remaining $m_B = N - m_A$ qubits contain more or less the same (classical) information. It is only for either a small number of qubits A or a large number of qubits A (close to N) that we see nonclassical behavior. For m_A close to N the mutual information is larger than 1 and thus contains quantum information in addition to the 1 bit of classical information, and we also see that that information is unique to A: both the mutual information and the unique information approach a value of 2 (and attain it for $m_A = N$).

We recall that the bonus $\mathcal{B}$ quantifies the asymmetry between A and B with respect to their correlations with $\mathcal{T}$. For $m_A \ll m_B$ that asymmetry is clearly large. This explains why the unique information in A is nonzero even when m_A is small. It may even be (slightly) larger than the mutual information $I(\mathcal{T};A)$, which, as mentioned at the end of Sec. II, can occur for classical probability distributions as well. In Fig. 4 this occurs only when A consists of just one qubit.

V. DISCUSSION AND CONCLUSION

We introduced a quantum version of the partial information decomposition, which defines synergistic, redundant, and unique information and which has proven to be a useful extension of Shannon's information theory. It allows one to distinguish correlations and entanglement between quantum systems that cannot be distinguished by standard versions of quantum mutual entropy. We applied the concept to quantum scrambling and quantified the idea that, roughly speaking, the no-cloning theorem forces there to be unique information. We saw this specifically in the Hayden-Preskill model of a black hole. We defined unique information in terms of a nonnegative quantity $\mathcal{B}_{\mathcal{Q}}$ [given by Eq. (21)] which is zero iff there is no unique information. The quantity $\mathcal{B}_{\mathcal{Q}}$ also measures the lack of symmetry between two systems with respect to their entanglement with a third reference system. That is,

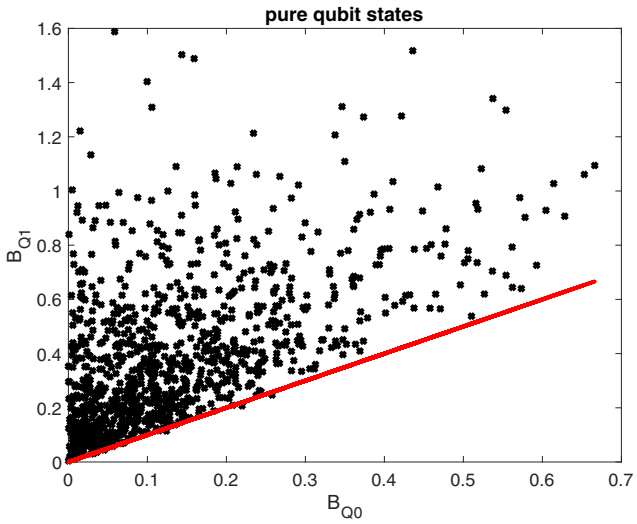

FIG. 6. For 1000 randomly chosen pure states of three qubits, $\mathcal{B}_{\mathcal{Q}1}$ vs $\mathcal{B}_{\mathcal{Q}0}$, with the red solid line indicating $\mathcal{B}_{\mathcal{Q}1} = \mathcal{B}_{\mathcal{Q}0}$. For slightly less than 1% of the states $\mathcal{B}_{\mathcal{Q}1}$ (just!) fails to be larger than $\mathcal{B}_{\mathcal{Q}0}$.

symmetry reduces unique information. We note that, likewise, symmetry reduces scrambling [54,55].

We end by suggesting three possible further applications. First, the point of quantum Darwinism is that classicality emerges from quantum mechanics when many copies of the same information exist in the environment [39,50,51,53]. One can quantify this notion in terms of unique information (or, rather, the lack of it) as an alternative to the standard analysis in terms of quantum discord [which is the difference between the (quantum) quantities I and J, as defined in Sec. III A]. Some preliminary results in this direction were given in Sec. IV C.

Second, a powerful way to analyze and characterize quantum chaos and, more generally, how quantum information propagates in quantum many-body systems is in terms of out-of-time-order correlators (OTOCs) [52,56–59]. These are observable quantities, but certain information measures, specifically Rényi entropies, can be measured as well [60,61]. Such information measures often appear as bounds on OTOCs. Thus, Rényi entropy versions of the quantities defined here may be useful in this context.

Third, given that the classical PID has been applied to understanding classical neural networks and given that ideas about scrambling can be applied to understanding or describing quantum neural networks [62–64], it seems that the quantum version of the PID could be fruitfully employed to understand quantum neural networks as well.

APPENDIX

We mention here some numerical results for random states (both pure and mixed) of three qubits and three qutrits. The point is to see how often $\mathcal{B}_{\mathcal{Q}1} > \mathcal{B}_{\mathcal{Q}0}$, that is, how often the nontrivial logarithmic pooling method beats the trivial method of choosing the lowest-entropy distribution.

For three-qubit states $\rho_{\mathcal{T}AB}$ we do find occasionally that the logarithmic pooling method is worse than simply choosing the lower- entropy distribution. This is true for a small fraction (11.4% from a sample of 10^5) of random mixed states and also for a smaller fraction of random pure states (0.83% from a sample of 10^5). See Figs. 5 and 6 for typical results for samples of size 1000.

For random three-qutrit states $\rho_{\mathcal{T}AB}$ we found numerically that the logarithmic pooling method almost always is superior to the simple method of choosing the lower-entropy distribution. Even for mixed states none of the sample of 10^5 states had $\mathcal{B}_{\mathcal{Q}0} > \mathcal{B}_{\mathcal{Q}1}$.

[1] P. L. Williams and R. D. Beer, Nonnegative decomposition of multivariate information, arXiv:1004.2515.

[2] For two variables Shannon's mutual information [30,31] suffices.

[3] R. G. James and J. P. Crutchfield, Multivariate dependence beyond Shannon information, Entropy **19**, 531 (2017).

[4] Coinformation is defined as $I(\mathcal{T};A,B) - I(\mathcal{T};A) - I(\mathcal{T};B)$. It is also known as interaction information [65] and has also been denoted as synergy [21–23]. In terms of the PID, coinformation equals the difference between synergistic and redundant information [1].

[5] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, M. M. Murray, T. P. Peixoto, F. Vaccarino, and G. Petri, The physics of higher-order interactions in complex systems, Nat. Phys. **17**, 1093 (2021).

[6] F. E. Rosas, P. A. M. Mediano, M. Gastpar, and H. J. Jensen, Quantifying high-order interdependencies via multivariate extensions of the mutual information, Phys. Rev. E **100**, 032305 (2019).

[7] T. M. S. Tax, P. A. M. Mediano, and M. Shanahan, The partial information decomposition of generative neural network models, Entropy **19**, 474 (2017).

[8] M. Wibral, C. Finn, P. Wollstadt, J. T. Lizier, and Viola Priesemann, Quantifying information modification in developing neural networks via partial information decomposition, Entropy **19**, 494 (2017).

[9] M. Wibral, V. Priesemann, J. W. Kay, J. T. Lizier, and W. A. Phillips, Partial information decomposition as a unified approach to the specification of neural goal functions, Brain Cognit. **112**, 25 (2017).

[10] E. L. Newman, T. F. Varley, V. K. Parakkattu, S. P. Sherrill, and J. M. Beggs, Revealing the dynamics of neural information processing with multivariate information decomposition, Entropy **24**, 930 (2022).

[11] J. T. Lizier, N. Bertschinger, J. Jost, and M. Wibral, Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work, Entropy **20**, 307 (2018).

[12] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay, Quantifying unique information, Entropy **16**, 2161 (2014).

[13] R. G. James, J. Emenheiser, and J. P. Crutchfield, Unique information and secret key agreement, Entropy **21**, 12 (2018).

[14] M. Harder, C. Salge, and D. Polani, Bivariate measure of redundant information, Phys. Rev. E **87**, 012130 (2013).

[15] V. Griffith and C. Koch, Quantifying synergistic mutual information, *Guided Self-Organization: Inception* (Springer, Berlin, Heidelberg, 2014), pp. 159–190.

[16] R. A. A. Ince, Measuring multivariate redundant information with pointwise common change in surprisal, Entropy **19**, 318 (2017).

[17] C. Finn and J. T. Lizier, Pointwise partial information decomposition using the specificity and ambiguity lattices, Entropy **20**, 297 (2018).

[18] A. Kolchinsky, A novel approach to the partial information decomposition, Entropy **24**, 403 (2022).

[19] F. E. Rosas, P. A. M. Mediano, B. Rassouli, and A. B. Barrett, An operational information decomposition via synergistic disclosure, J. Phys. A **53**, 485001 (2020).

[20] S. J. van Enk, Pooling probability distributions and partial information decomposition, Phys. Rev. E **107**, 054133 (2023).

[21] N. Brenner, S. P. Strong, R. Koberle, W. Bialek, and R. R. de Ruyter van Steveninck, Synergy in a neural code, Neural Comput. **12**, 1531 (2000).

[22] E. Schneidman, II, W. Bialek, and M. J. Berry, Synergy, redundancy, and independence in population codes, J. Neurosci. **23**, 11539 (2003).

[23] P. E. Latham and S. Nirenberg, Synergy, redundancy, and independence in population codes, revisited, J. Neurosci. **25**, 5195 (2005).

[24] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[25] M. Hayashi, *Quantum Information Theory* (Springer, Berlin, Heidelberg, 2016).

[26] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2013).

[27] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).

[28] P. Hayden and J. Preskill, Black holes as mirrors: Quantum information in random subsystems, J. High Energy Phys. 09 (2007) 120.

[29] H. S. Dhar, A. K. Pal, D. Rakshit, A. Sen, and U. Sen, Monogamy of quantum correlations-a review, in *Lectures on General Quantum Correlations and Their Applications* (Spring Nature, Cham, Switzerland, 2017), pp. 23–64.

[30] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[31] T. M. Cover, *Elements of Information Theory* (Wiley, New York, 1999).

[32] A. Rényi, On measures of entropy and information, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1:, Contributions to the Theory of Statistics* (University of California Press, Berkeley, 1961), Vol. 4, pp. 547–562.

[33] For example, Ref. [13] defines a PID in terms of secret key rates. One expects ideas from quantum secret sharing [66] to lead to a QPID, too.

[34] A. Bhattacharyya, On a measure of divergence between two multinomial populations, Sankhyā: Indian J. Stat. **7**, 401 (1946), http://www.jstor.org/stable/25047882.

[35] N. C. Lind and A. S. Nowak, Pooling expert opinions on probability distributions, J. Eng. Mech. **114**, 328 (1988).

[36] D. Poole and A. E. Raftery, Inference for deterministic simulation models: The Bayesian melding approach, J. Am. Stat. Assoc. **95**, 1244 (2000).

[37] L. M. Carvalho, D. A. M. Villela, F. C. Coelho, and L. S. Bastos, Bayesian inference for the weights in logarithmic pooling, Bayesian Anal. **18**, 223 (2023).

[38] E. Neyman and T. Roughgarden, No-regret learning with unbounded losses: The case of logarithmic pooling, arXiv:2202.11219.

[39] H. Ollivier and W. H. Zurek, Quantum discord: A measure of the quantumness of correlations, Phys. Rev. Lett. **88**, 017901 (2001).

[40] Y. Huang, Computing quantum discord is np-complete, New J. Phys. **16**, 033027 (2014).

[41] C. H. Bennett and S. J. Wiesner, Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states, Phys. Rev. Lett. **69**, 2881 (1992).

[42] Z. Zhang and S. Luo, Probabilistic interpretation of the reduction criterion for entanglement, Phys. Rev. A **75**, 032312 (2007).

[43] M. S. Leifer and R. W. Spekkens, Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference, Phys. Rev. A **88**, 052130 (2013).

[44] M. S. Leifer and R. W. Spekkens, A Bayesian approach to compatibility, improvement, and pooling of quantum states, J. Phys. A **47**, 275301 (2014).

[45] N. J. Cerf and C. Adami, Negative entropy and information in quantum mechanics, Phys. Rev. Lett. **79**, 5194 (1997).

[46] N. J. Cerf and C. Adami, Quantum extension of conditional probability, Phys. Rev. A **60**, 893 (1999).

[47] M. Vempati, N. Ganguly, I. Chakrabarty, and A. K. Pati, Witnessing negative conditional entropy, Phys. Rev. A **104**, 012417 (2021).

[48] N. Friis, S. Bulusu, and R. A. Bertlmann, Geometry of two-qubit states with negative conditional entropy, J. Phys. A **50**, 125301 (2017).

[49] We could trivially generalize P_1 and P_2 to mixed states ρ_1 and ρ_2, and then $\mathcal{B}_\mathcal{Q}$ simply reduces to $\mathcal{B}$ by construction.

[50] M. Zwolak and W. H. Zurek, Complementarity of quantum discord and classically accessible information, Sci. Rep. **3**, 1729 (2013).

[51] T. K. Unden, D. Louzon, M. Zwolak, W. H. Zurek, and F. Jelezko, Revealing the emergence of classicality using nitrogen-vacancy centers, Phys. Rev. Lett. **123**, 140402 (2019).

[52] A. Touil and S. Deffner, Quantum scrambling and the growth of mutual information, Quantum Sci. Technol. **5**, 035005 (2020).

[53] D. Girolami, A. Touil, B. Yan, S. Deffner, and W. H. Zurek, Redundantly amplified information suppresses quantum correlations in many-body systems, Phys. Rev. Lett. **129**, 010401 (2022).

[54] H. Tajima and K. Saito, Universal limitation of quantum information recovery: Symmetry versus coherence, arXiv:2103.01876.

[55] J. Kudler-Flam, R. Sohal, and L. Nie, Information scrambling with conservation laws, SciPost Phys. **12**, 117 (2022).

[56] P. Hosur, X.-L. Qi, D. A. Roberts, and B. Yoshida, Chaos in quantum channels, J. High Energy Phys. 02 (2016) 004.

[57] D. Ding, P. Hayden, and M. Walter, Conditional mutual information of bipartite unitaries and scrambling, J. High Energy Phys. 02 (2016) 145.

[58] E. Iyoda and T. Sagawa, Scrambling of quantum information in quantum many-body systems, Phys. Rev. A **97**, 042330 (2018).

[59] S. Xu and B. Swingle, Scrambling dynamics and out-of-time ordered correlators in quantum many-body systems: A tutorial, arXiv:2202.07060.

[60] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, Science **364**, 260 (2019).

[61] S. J. van Enk and C. W. J. Beenakker, Measuring Tr ρ^n on single copies of ρ using random measurements, Phys. Rev. Lett. **108**, 110503 (2012).

[62] H. Shen, P. Zhang, Y.-Z. You, and H. Zhai, Information scrambling in quantum neural networks, Phys. Rev. Lett. **124**, 200504 (2020).

[63] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021).

[64] M. Sajjan, V. Singh, R. Selvarajan, and S. Kais, Imaginary components of out-of-time-order correlator and information scrambling for navigating the learning landscape of a quantum machine learning model, Phys. Rev. Res. **5**, 013146 (2023).

[65] W. McGill, Multivariate information transmission, Trans. IRE Prof. Group Inf. Theory **4**, 93 (1954).

[66] M. Hillery, V. Bužek, and A. Berthiaume, Quantum secret sharing, Phys. Rev. A **59**, 1829 (1999).